MOOMIN – Mathematical explOration of 'Omics data on a MetabolIc Network

Taneli Pusa, Mariana G. Ferrarini, Ricardo Andrade, Arnaud Mary, Alberto Marchetti-Spaccamela, Leen Stougie, Marie-France Sagot

Bioinformatics, 2019, 1–10
doi: 10.1093/bioinformatics/btz584
Advance Access Publication Date: 22 August 2019
Original Paper

Gene expression

MOOMIN – Mathematical explOration of 'Omics data on a Metabollc Network

Taneli Pusa[1,2,3,*], Mariana Galvão Ferrarini [2,4], Ricardo Andrade[1,2], Arnaud Mary[1,2], Alberto Marchetti-Spaccamela[3], Leen Stougie[5] and Marie-France Sagot[1,2,*]

[1]INRIA Grenoble Rhône-Alpes, Montbonnot-Saint-Martin 38334, France, [2]Laboratoire de Biométrie et Biologie Évolutive, UMR 5558, CNRS, Université de Lyon, Université Lyon 1, Villeurbanne 69622, France, [3]Department of Computer, Automatic and Management Engineering, Sapienza University of Rome, Rome 00185, Italy, [4]Univ Lyon, INSA-Lyon, INRA, BF2i, UMR0203, F-69621, Villeurbanne 69622, France and [5]CWI, 1098XG Amsterdam, The Netherlands

*To whom correspondence should be addressed.

Associate Editor: Lenore Cowen

Received on February 28, 2019; revised on July 16, 2019; editorial decision on July 18, 2019; accepted on August 19, 2019

Abstract

Motivation: Analysis of differential expression of genes is often performed to understand how the metabolic activity of an organism is impacted by a perturbation. However, because the system of metabolic regulation is complex and all changes are not directly reflected in the expression levels, interpreting these data can be difficult.

Results: In this work, we present a new algorithm and computational tool that uses a genome-scale metabolic reconstruction to infer metabolic changes from differential expression data. Using the framework of constraint-based analysis, our method produces a qualitative hypothesis of a change in metabolic activity. In other words, each reaction of the network is inferred to have increased, decreased, or remained unchanged in flux. In contrast to similar previous approaches, our method does not require a biological objective function and does not assign on/off activity states to genes. An implementation is provided and it is available online. We apply the method to three published datasets to show that it successfully accomplishes its two main goals: confirming or rejecting metabolic changes suggested by differentially expressed genes based on how well they fit in as parts of a coordinated metabolic change, as well as inferring changes in reactions whose genes did not undergo differential expression.

Availability and implementation: github.com/htpusa/moomin.

Contact: henri.pusa@inria.fr or marie-france.sagot@inria.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

In recent decades, the advent of high-throughput sequencing has greatly increased the amount of biological data available. Transcriptomics, that is, the sequencing of mRNA, offers a cost-efficient way to quantify the expression of genes within a cell. Specifically RNA-Seq, which is able to measure in absolute terms the amount of expression across the whole genome, is now becoming standard practice.

However, with increasing data comes an increasing need for apt analytic pipelines. Especially in the area of differential expression (DE) analysis, where the expression of genes is compared between two or more different conditions, there is still much to be done to fully actualize the potential of 'omics data. While specialized statistical methodology has been developed and adopted, it can only provide a list of genes that have passed the test for significant change. Due to the complex nature of regulatory actions that take place after

transcription, this information alone is often not as enlightening as one would hope. Answering questions such as 'what metabolic reorganization followed from an external change' based on expression levels alone remains thus difficult.

Where transcriptomics data serves as a hint as to the actual functioning of a cell, the metabolic network represents the catalogue of possible functions and their relationships. It is then a natural choice to combine gene expression data with the knowledge provided by a metabolic network.

The simplest representation of a metabolic network is by means of an ordinary graph that connects together either reactions or their associated genes based on shared metabolites. Because these connections do not necessarily correspond to a flow of matter, such simple model is perhaps best described as a knowledge graph where edges represent associations rather than physical structures. The earliest approaches to incorporating the metabolic network into the analysis of 'omics data thus correspond to searching for subsets of interest amongst metabolic genes or reactions by way of subgraph extraction: starting from a set of nodes of interest (i.e. genes or reactions implicated by the data), find a subgraph that connects the observations (Alcaraz *et al.*, 2011, 2012; Antonov *et al.*, 2008; Baumbach *et al.*, 2012; Noirel *et al.*, 2008; van Helden *et al.*, 2000). Similar methods have also been used with other networks, such as protein–protein (Segal *et al.*, 2003) or gene interaction networks (Ideker *et al.*, 2002).

A richer way of representing and analyzing metabolic networks is through the framework of constraint-based analysis. Here, the network is given as a stoichiometric matrix that not only contains a more accurate graphical representation of the network structure, but also includes the reaction stoichiometry. The most notable application of this formalism is flux balance analysis (FBA, Orth *et al.*, 2010). In FBA, the goal is to predict intracellular fluxes based on reaction stoichiometry, an assumption of a steady state where the accumulation of intracellular metabolites is forbidden, and some biological objective function. Gene expression data, which in principle should provide evidence about the abundance of catalytic enzymes, can be used to further constrain the space of possible fluxes, either by switching off reactions for which the associated genes were not detected (Akesson *et al.*, 2004; Becker and Palsson, 2008; Chandrasekaran and Price, 2010; Collins *et al.*, 2012; Covert and Palsson, 2003) or by using the amount of mRNA as a proxy for the enzymatic activity (Colijn *et al.*, 2009; Kim *et al.*, 2013, 2016; Lee *et al.*, 2012; Schwarz *et al.*, 2005; Tian and Reed, 2018; Zhang *et al.*, 2017). In the context of DE data, instead of absolute gene expression levels, changes in expression have been used to predict fluxes in a perturbed state (Fang *et al.*, 2012; Moxley *et al.*, 2009).

Constraint-based methods have been used in conjunction with 'omics data to build so-called context-specific models: metabolic reconstructions appropriately pruned to correspond to the active part of metabolism in a given tissue or environment (Agren *et al.*, 2012; Jerby *et al.*, 2010; Robaina Estevez and Nikoloski, 2015; Rossell *et al.*, 2013; Schultz and Qutub, 2016; Shlomi *et al.*, 2008; Vlassis *et al.*, 2014; Wang *et al.*, 2012; Yizhak *et al.*, 2014). Similarly to gene expression guided FBA, reaction activity is first constrained using gene expression data and reactions that are predicted to be inactive are removed.

However, surprisingly little research has been done at the intersection of constraint-based analysis and DE. Because the framework behind constraint-based methods directly describes the flow of matter through the network in terms of reaction fluxes, it can be used to generate explicit hypotheses of metabolic shifts from expression data.

We briefly describe three works that in our view, are more relevant in this context. In the first, Jensen and Papin (2011) described MADE, a method that takes as its input a time-series of gene expression and tries to find a series of binary activity states for each reaction. Transitions from active to inactive and *vice versa* are scored based on DE analysis performed comparing each time point to the previous one so that differences that correspond to a statistically significant change in expression are promoted. In another work, Samal *et al.* (2017) used extreme currents (EC, Clarke, 1988) to find sets of genes whose activity correlated with a certain condition. The authors first calculated the ECs of a network leading to sets of genes. Then, given a gene expression dataset spanning several conditions, a statistical model was fitted to find sets of genes associated with certain conditions. Finally, Zhu *et al.* (2017) first defined a set of reactions that changed in expression from one condition to another by looking for changes in the average expression levels of all associated genes. They then formulated a MILP-problem which minimizes an inconsistency score between two flux vectors, one for each condition and the observed changes in expression levels. For a review of combining 'omics data and metabolic networks, we refer the reader to (Kim and Lun, 2014; Machado and Herrgård, 2014; Opdam *et al.*, 2017; Rezola *et al.*, 2015; Saha *et al.*, 2014; Vivek-Ananth and Samal, 2016).

In this paper, we propose a new method called `MOOMIN`, for Mathematical explOration of 'Omics data on a MetabolIc Network, that we believe offers several improvements in the integration of transcriptomic data with metabolic networks in relation to previous methods. The first improvement is that our method does not require a biological objective function (as opposed to, for example, Jensen and Papin, 2011), which sometimes might be problematic. Take the biomass objective function for instance, which corresponds to the production of biomass as an approximation of growth: frequently this function is not tailor-made for the organism in question, but is rather a reproduction of objective functions for model species such as *Escherichia coli* or human. As it is not entirely clear how an imprecise biological objective function may influence the final results, we did not want to rely on any potentially inaccurate information.

Secondly, we do not use information of pre-defined metabolic pathways to validate or discard reactions. This information is still available and can be used *a posteriori* to interpret the results. In this sense, the method might be able to discover new pathways, which should be evaluated by the user as biologically meaningful or not. Another advantage offered by this aspect of our approach is that the obtained solutions do not only indicate distinct pathways as up- or down-regulated, but also show how these changes are connected. Moreover, similarly to Jensen and Papin (2011), Zhu *et al.* (2017) and opposed to for example Zur *et al.* (2010), we do not use absolute thresholds to define the activity of genes. However, we also do not require genes to have a binary inactive/active state, as was proposed by Jensen and Papin (2011).

Thirdly, to our knowledge all previous methods are designed with frequentist statistics in mind. This in practice means that changes in expression are detected using the *P*-value. However, the *P*-value is not suited for detection of the absence of a change. Our third improvement is thus that `MOOMIN` uses the Bayesian posterior probability of differential expression, allowing us to identify not only genes that changed in expression, but also those that did not. While previous methods have used the *P*-value to this end (Jensen and Papin, 2011), we argue that by instead quantifying our beliefs about the observed changes in terms of posterior probabilities is theoretically sounder. We shall therefore present in the next sections the method we developed and the approaches we chose to improve the state-of-the-art algorithms as well as two case-studies where our method was able to detect results that were not evident from the differential expression data.

Finally, we provide an implementation as a collection of MATLAB scripts to be used in concert with the COBRA Toolbox (Heirendt *et al.*, 2019). This is particularly interesting because the closest approach in terms of methodology (Zhu *et al.*, 2017) does not have an implementation available, as far as we are aware. Moreover, the method of Zhu *et al.* does not take as input the results of a DE analysis but rather the authors have created their own pipeline that directly identifies changes at the level of the reactions. This is a rather non-standard approach and it is unclear how exactly the average expression level behaves as an indicator of change in activity. Because differential expression is determined using a t-test, only those reactions that underwent a change can be identified.

2 Approach

We propose a method to combine the information from a RNA-seq DE analysis with a genome-scale metabolic network to infer which parts of an organism's metabolism increased or decreased in activity. To this end we introduce the concept of a *feasible change*: any inferred change should be such that it corresponds to a transition from one steady state (zero accumulation for all internal metabolites) to another.

The change in metabolic activity is expressed in qualitative terms as *colours*. Namely, we colour reactions red (resp. blue) to signify an increase (resp. decrease) in metabolic flux.

The gene expression information is leveraged by turning the results of a DE analysis into *weights* for the reactions of the network: evidence for a positive or negative change in expression will result in a positive weight that promotes the inclusion of said reaction in the solution. Conversely, when there is little or no evidence of change, a negative weight is given to discourage including the reaction. The weights are calculated based on the strength of the evidence for a change in expression. An optimal solution is one that maximizes the sum of the weights of the included reactions.

The feasibility of the change is ensured using stoichiometric constraints. We also show how to approximate these constraints using the topology of the network, represented not as a simple graph but as a hypergraph (Klamt *et al.*, 2009; Lacroix *et al.*, 2008), to decrease the running time of the algorithm.

The optimization problem is formulated using mixed-integer linear programming (MILP) and can be solved by any specialized solver. In addition to finding a single optimal solution, we can also enumerate all optimal solutions.

MOOMIN was implemented in MATLAB and relies on the COBRA Toolbox (Heirendt *et al.*, 2019). A MILP-solver compatible with the COBRA Toolbox is also required. The code is available for download at github.com/htpusa/moomin.

3 Materials and methods

We are studying a change in the state of metabolism from one condition to another. Let us call these two states 1 and 2. They can be, for example, two different nutrient environments, or state 2 could be a knock-out where state 1 is the wild-type.

In the classic FBA approach, metabolism is assumed to be at a steady state, that is,

$$S \cdot \mathbf{v}_1 = 0 \quad \text{and} \quad S \cdot \mathbf{v}_2 = 0 \tag{1}$$

where S is the stoichiometric matrix, and $\mathbf{v}_1$ and $\mathbf{v}_2$ are the flux vectors in the two conditions respectively. We wish to infer the change $\Delta\mathbf{v} = \mathbf{v}_2 - \mathbf{v}_1$.

We will make use of the fact that the change $\Delta\mathbf{v}$ is also a valid flux vector. In other words, any change from one steady state to another must also satisfy

$$S \cdot \Delta\mathbf{v} = 0 \tag{2}$$

However, because we do not believe it is possible to exactly quantify these changes based on RNA levels, and because we wish to avoid having to specify the two flux vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we shall concern ourselves only with qualitative changes.

Let $\mathbf{x}_i^+$ and $\mathbf{x}_i^-$ be binary variables that express, respectively, reaction i increasing or decreasing in flux from condition 1 to condition 2. In other words, $\mathbf{x}_i^+ = 1$ ($\mathbf{x}_i^- = 1$) implies that $\Delta\mathbf{v}_i > 0$ ($\Delta\mathbf{v}_i < 0$). This can be expressed as the following set of LP-constraints:

$$\Delta\mathbf{v}_i + \mathbf{x}_i^+\mathbf{l}_i > \mathbf{l}_i, \quad i = 1, \ldots, m \tag{3}$$

$$\Delta\mathbf{v}_i - \mathbf{x}_i^+\mathbf{u}_i \leq 0, \quad i = 1, \ldots, m \tag{4}$$

$$\Delta\mathbf{v}_i + \mathbf{x}_i^-\mathbf{u}_i < \mathbf{u}_i, \quad i = 1, \ldots, m \tag{5}$$

$$\Delta\mathbf{v}_i + \mathbf{x}_i^-(-\mathbf{l}_i) \geq 0, \quad i = 1, \ldots, m \tag{6}$$

where $\mathbf{u}$ and $\mathbf{l}$ are vectors of upper and lower bounds set for the change $\Delta\mathbf{v}$ (for a further explanation, see below). Note that in practice the strict inequalities in constraints (3) and (5) need to be implemented using a small threshold constant (see Supplementary Materials).

Together, the constraints in Equations (2) and (3)–(6) define the set of qualitative changes expressed by a pair of binary vectors $(\mathbf{x}^+, \mathbf{x}^-)$ that are stoichiometrically feasible.

We now wish to incorporate the gene expression data in order to find the changes that are also supported by observations. We do this by using the results of the DE analysis to formulate an objective function. We also impose constraints ruling out some changes that contradict the observations.

Let **PPDE** and **FC** be two data vectors such that $\mathbf{PPDE}_j$ is the *a posteriori* probability that gene j was differentially expressed between conditions 1 and 2, and $\mathbf{FC}_j$ is the associated $\log_2$ fold change observed. We call gene j *red* if $\mathbf{PPDE}_j > t$ and $\mathbf{FC}_j > 0$, where t is a chosen threshold of differential expression (the value 0.9 is used throughout this paper). A gene is *blue* if $\mathbf{PPDE}_j > t$ and $\mathbf{FC}_j < 0$, and finally *grey* if $\mathbf{PPDE}_j < t$, that is, it was not differentially expressed. We transform the probabilities to weights using the function:

$$w_j = \min\{\beta(-\log(1 - \mathbf{PPDE}_j) + \log(1 - t)), -\alpha\beta\log(1 - t)\} \tag{7}$$

where α and β are parameters. In this way, red and blue genes have a positive weight, and grey genes a negative one, proportional to their probability of having been differentially expressed.

Every reaction of the metabolic network has a set of genes associated with it (so-called gene-protein-reaction association or GPR). We give a colour and a weight to each reaction based on these associated genes according to the following rules. If at least one of the associated genes is red (blue), and no other gene is blue (red), we colour the reaction red (blue). If there are no red or blue genes, the reaction is coloured grey. In all of these cases, the weight of the reaction will be the maximum of the weights of the associated genes.

If there is a contradiction, that is, if there is at least one red and one blue gene, we colour the reaction grey and give it a weight corresponding to that of a gene with the probability of being differentially expressed 0.5. Finally, if the set of genes is empty, that is, we do not

know of any genes associated with the reaction, we colour the reaction grey and give it a weight corresponding to that of a gene with the probability of differential expression 0.

We denote these *a priori* colours, at this point corresponding only to proposed changes and not necessarily globally feasible, by a pair of binary vectors with a special subscript $(\mathbf{x}_0^+, \mathbf{x}_0^-)$.

We treat the colours given based on the data merely as suggestions, and thus as the starting point of the algorithm. They can still be rejected during the final search. In other words, *a priori* coloured reactions can appear as grey in a solution, and grey reactions can become coloured. However, we do exclude solutions that go completely against the findings of the differential expression analysis. More precisely, if a reaction is coloured red by its genes, it cannot be blue in the solution, and *vice versa*.

Finding the feasible change that best fits the data can now be expressed as the following MILP-problem

$$\underset{(\mathbf{x}^+,\mathbf{x}^-)}{\text{maximise}} \quad \sum_{i=1}^{m}(\mathbf{x}_i^+ + \mathbf{x}_i^-)w_i \tag{8}$$

$$\text{s.t.} \quad \mathbf{x}^+{}_i + \mathbf{x}^-{}_i \leq 1, \quad i = 1, \ldots, \mathrm{m} \tag{9}$$

$$\mathbf{x}_i^+ = 0 \quad \forall i \quad \text{s.t.} \quad \mathbf{x}_{0\,i}^- = 1 \quad \text{and} \quad \mathbf{u}_i = 0 \tag{10}$$

$$\mathbf{x}_i^- = 0 \quad \forall i \quad \text{s.t.} \quad \mathbf{x}_{0\,i}^+ = 1 \quad \text{and} \quad \mathbf{l}_i = 0 \tag{11}$$

$$S \cdot \mathbf{v} = 0 \tag{12}$$

$$\mathbf{v}_i + \mathbf{x}_i^+\mathbf{l}_i > \mathbf{l}_i, \qquad i = 1, \ldots, m \tag{13}$$

$$\mathbf{v}_i - \mathbf{x}_i^+\mathbf{u}_i \leq 0, \qquad i = 1, \ldots, m \tag{14}$$

$$\mathbf{v}_i + \mathbf{x}_i^-\mathbf{u}_i < \mathbf{u}_i, \qquad i = 1, \ldots, m \tag{15}$$

$$\mathbf{v}_i + \mathbf{x}_i^-(-\mathbf{l}_i) \geq 0, \qquad i = 1, \ldots, m \tag{16}$$

$$\mathbf{l}_i \leq \mathbf{v}_i \leq \mathbf{u}_i, \qquad i = 1, \ldots, m \tag{17}$$

$$\mathbf{x}^+, \mathbf{x}^- \in \{0,1\}^m \tag{18}$$

$$\mathbf{v} \in \mathbb{R}^m \tag{19}$$

We have shortened $\Delta\mathbf{v}$ above to $\mathbf{v}$ to simplify the notation. It should be stressed that it is not a flux vector in the normal sense: because it represents a difference between two steady state flux distributions, entries can be negative even for irreversible reactions. Its magnitude is also not important since we are interested only in the qualitative nature of the change. Thus the flux bounds $\mathbf{u}$ and $\mathbf{l}$ also serve a different role: in general, we can simply set $\mathbf{u}_i = c$ and $\mathbf{l}_i = -c$ for all i, where c is some positive constant. The exception are the *a priori* coloured reactions. As described before, we forbid *a priori* red (blue) reactions from becoming blue (red). This is done by simply setting $\mathbf{u}_i = 0$ ($\mathbf{l}_i = 0$) for *a priori* blue (red) reactions to prevent them from being coloured against the *a priori* colour.

For reversible reactions the enforcement of the *a priori* colours needs to be done differently. This is because an increase in flux from 1 to 2 in the primary direction ($\mathbf{v}_{1i} > 0$ and $\mathbf{v}_{2i} > 0$) is equivalent in our formulation to a decrease in flux in the reverse direction ($\mathbf{v}_{1i} < 0$ and $\mathbf{v}_{2i} < 0$) since in both cases $\Delta\mathbf{v}_i > 0$. Consider the situation depicted in Figure 1. Above, the *a priori* colouring shows an increase in the flux of reaction r_1. For the metabolite balance this means that the net production of metabolite B is increased. Let us

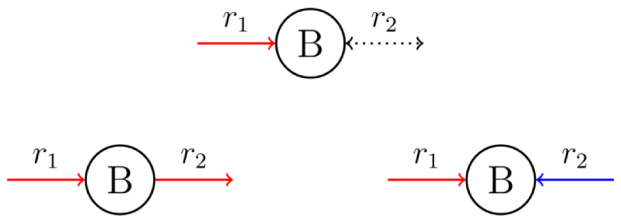

Fig. 1. Reversible reactions can have two different assumptions of a shift that result in the same changes in metabolite balances. The increase in flux in r_1 implies that either more of B should be consumed or less produced in reaction r_2, depending on its direction. Both of these scenarios correspond to $\Delta\mathbf{v}_2 > 0$ but the colour given to r_2, either red for an increase (left) or blue for a decrease in flux (right), depends on the interpretation

assume that the 'default' direction of r_2 is to consume B, that is, a flux in this direction would be positive ($\mathbf{v}_{12} > 0$ and $\mathbf{v}_{22} > 0$). The increased production of B by r_1 can be balanced by assuming an increase in r_2 as well, that is, colouring it red. This corresponds to assuming that $\Delta\mathbf{v}_2 > 0$. However, $\Delta\mathbf{v}_2 > 0$ is also consistent with $\mathbf{v}_{12} < 0$ and $\mathbf{v}_{22} < 0$, that is, r_2 being in the 'reverse' direction, and colouring it blue for a *decrease* in flux. For *a priori* coloured, reversible reactions this means that their flux bounds remain unaffected since the $\Delta\mathbf{v}_i$ can be assigned seemingly 'against' the *a priori* colour, and simply leads to the conclusion that the direction of the reaction is the reverse one. Unfortunately this also means that the colour and direction of *a priori* grey, reversible reactions can remain ambiguous. However, we expect that in most cases the surrounding network can be used to determine the correct interpretation. When necessary, we have used the colour yellow for such reactions.

An optimal solution is not necessarily unique, and it can be of interest to look for alternative optima to determine if they contain alternative biological hypotheses. Enumeration of all optimal solutions can be done iteratively by adding constraints to exclude already discovered ones. In order to reduce the number of possible alternative optima, we only look for alternative solutions in terms of which reactions are coloured, and not the colours themselves. Let Z_{opt} be the optimal value of the objective function and $\mathbf{x}_k \in \mathbb{R}^m$ the kth solution obtained, where $\mathbf{x}_{kj} = \mathbf{x}_{kj}^+ \wedge \mathbf{x}_{kj}^-$, that is, it is a vector indicating if a reaction is coloured. To obtain $\mathbf{x}_{k+1}$, we add the following constraints to the original MILP problem:

$$\sum_{i=1}^{m}(\mathbf{x}_i^+ + \mathbf{x}_i^-)w_i = Z_{opt} \tag{20}$$

$$2(\mathbf{x}_k \cdot \mathbf{x}_{k+1}) \leq \sum_i^m \mathbf{x}_{ki} + \sum_i^m \mathbf{x}_{k+1i} - 1 \tag{21}$$

While in our experiments we were able to solve the optimization problem in Equation (8) for genome-scale networks of prokaryotes in the matter of minutes using a standard laptop computer, larger networks and the enumeration of multiple optima can result in impractical running times. To account for this, the implementation of `MOOMIN` contains an alternative problem formulation that uses simpler topological constraints to enforce the feasibility of the inferred shift. This 'topological formulation' often results in an optimization problem that is much easier to solve, yet provides solutions that largely overlap with those of the original problem. However, since it does not guarantee full stoichiometric balance, artifacts such as infeasible cycles can be produced. The topological formulation is described in more detail in the Supplementary Materials.

4 Results

To test and demonstrate MOOMIN, we chose three previously published DE datasets. For all three sets, we started the analysis from raw counts and performed the DE analysis using the R package EBSeq (Leng *et al.*, 2013). For the last two datasets, we also performed a standard gene ontology (GO) enrichment analysis using the EBSeq output. The results are shown in Table 1 and demonstrate some differences between 'traditional' DE analysis and our algorithm (for full results of the GO enrichment, see Supplementary Tables).

The first dataset analyzed compared *Saccharomyces cerevisiae* growth in a chemostat versus a batch culture, and is meant as a proof of concept as the associated metabolic shift is both well understood and clearly reflected in the expression of genes. These results are discussed in Section 4.1. The other two datasets studied *Escherichia coli* exposure to mercury and *S.cerevisiae* exposure to ethanol, and offer more biologically interesting results. They are discussed in Sections 4.2 and 4.3 respectively.

For all datasets, we also compared the outputs from the original formulation [Equation (8)] and the topological approximation (see Supplementary Materials). These were largely in agreement, with only minor differences in reaction colours. A closer look at how the two solutions might diverge is provided for the *S.cerevisiae* ethanol stress data in Section 4.3.1.

The *S.cerevisiae* model used was the iMM904 published by Mo *et al.* (2009). It contains 1226 metabolites and 1577 reactions. The *E.coli* model chosen was the iJO1366 published by Orth *et al.* (2011). It contains 1805 metabolites and 2583 reactions. Both models were downloaded from the BiGG database (King *et al.*, 2016). MOOMIN was run on an early 2015 MacBook Pro with a two-core 2, 9 GHz Intel Core i5 processor, and solving the problem in Equation (8) took 446 s for the *S.cerevisiae* chemostat versus batch dataset, 72 s for the *E.coli* mercury exposure set, and 5182 s for the *S.cerevisiae* ethanol exposure set.

4.1 *Saccharomyces cerevisiae* growth in chemostat versus batch

This dataset was originally published by Nookaew *et al.* (2012) and contains gene expression measurements of the yeast *S.cerevisiae* obtained using RNA-Seq from two conditions: growth in a batch culture and growth in a chemostat culture. In a batch culture, a small number of cells are inoculated into a nutrient-rich medium, and allowed to grow at maximal rate until the nutrients are exhausted and the growth curve achieves saturation. In contrast, in a chemostat culture, nutrients are added to the culture at a fixed flow rate while the biomass and the products of metabolism are removed from the vessel at the same flow rate to maintain a fixed culture volume. Growth will settle to a steady state value and cells will continue to divide indefinitely.

As a result, and as already reported by several studies, up-regulated pathways in a chemostat culture (if we consider batch growth as the control) are usually involved with electron transport, aerobic respiration and the TCA cycle, which were reported by Nookaew *et al.* (2012), and also recovered by MOOMIN (12 out of 16 reactions for aerobic respiration and 11 out of 13 for the TCA cycle). Figure 2 shows a visualization of the MOOMIN output for the TCA cycle.

4.2 *Escherichia coli* mercury exposure

The second dataset analyzed, originally published by LaVoie and Summers (2018), demonstrates *E.coli*'s response to mercury exposure. Mercury is a toxicant that negatively impacts the health of both microscopic and macroscopic organisms, and induces a broad cellular response in *E.coli*. Gene expression was measured using RNA-Seq in two conditions: an unexposed control and a culture exposed to mercuric chloride ($HgCl_2$).

For this dataset, the optimum of Equation (8) was not unique, and in total, there were 1684 alternative optima. However, across all these solutions, 1120 of 2578 reactions were present as coloured in all solutions, and 1338 in none. Accordingly, analysis of the results was done using a consensus across all optima which for most reactions uniquely determined their colour.

MOOMIN was able to detect changes in all the pathways that showed up as differentially expressed in our GO analysis (Table 1) and were described by LaVoie and Summers (2018) as up- or down-regulated. As expected, reactions from amino acid and nucleotide biosynthesis, NADH metabolism, carbohydrate metabolism and glycolysis, whose corresponding genes were down-regulated, appeared with predicted reductions in flux in both topological and stoichiometric solutions (Supplementary Tables). Further in agreement with

Table 1. Pathways discovered using GO enrichment versus MOOMIN

	DE	Metabolic process[a]	GO enrichment	MOOMIN
E.coli mercury exposure	Up-regulated	Stress response	×	×
		Redox response	×	×
		Gluthatione and cysteine metabolism		×
		GABA shunt		×
	Down-regulated	Amino acid metabolism	×	×
		Nucleotiode salvage pathway	×	×
		NADH metabolism	×	×
		Carbohydrate metabolism	×	×
		Cell envelope biosynthesis		×
S.cerevisiae ethanol stress	Up-regulated	TCA cycle and anaplerotic reactions		×
		Arginine metabolism		×
		Organelle metabolism	×	×
	Down-regulated	Amino acid metabolism	×	×
		Alternate carbon metabolism	×	×
		Trehalose availability		×
		Fatty acid metabolism		×

[a]Not every GO term is directly related to metabolism, in this way, some terms present in Supplementary Tables might not be present here.

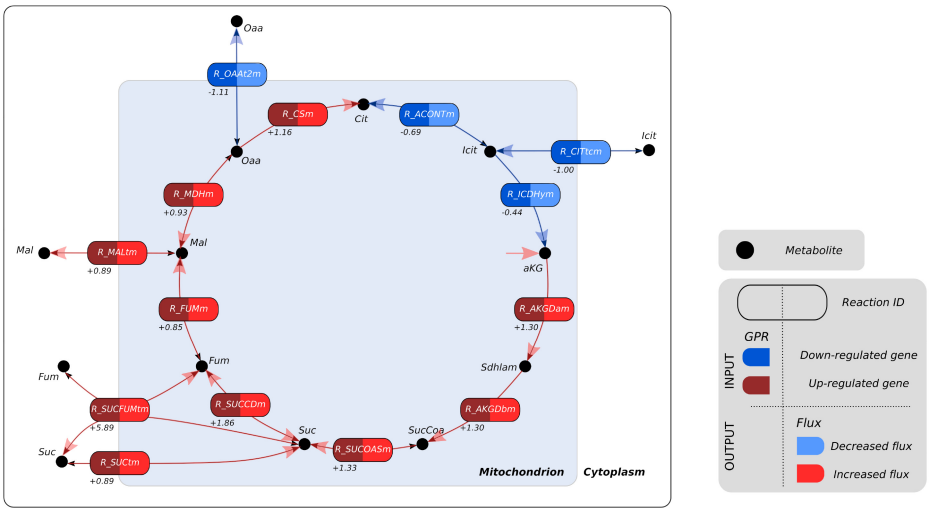

Fig. 2. An example of the predicted flux changes for the TCA cycle of yeast growth in chemostat mode versus batch mode. Round rectangles represent reactions whereas circles represent metabolites. Reaction rectangles are divided into the *a priori* (INPUT) and solution (OUTPUT) colours. The INPUT is dictated by the gene–protein-reaction (GPR) and the DE data as explained in Section 3. Below the INPUT is shown the $\log_2$ fold change of the differentially expressed gene responsible for the *a priori* colour. Here, we tested a proof of concept dataset of yeast growth in chemostat mode versus batch mode. The shift in metabolism between the two types of growth is known to affect the entire TCA cycle, as it is reflected by the up and down-regulated genes seen in dark red and dark blue (INPUT), respectively. In this example, all reactions had at least one differentially expressed gene and are thus *a priori* coloured. `MOOMIN` then confirms these *a priori* colours as well as chooses a direction for reversible reactions so that stoichiometric balance is maintained. For examples of `MOOMIN` also colouring *a priori* grey reactions (see Figs 3 and 4). The metabolite abbreviations are defined in Supplementary Tables

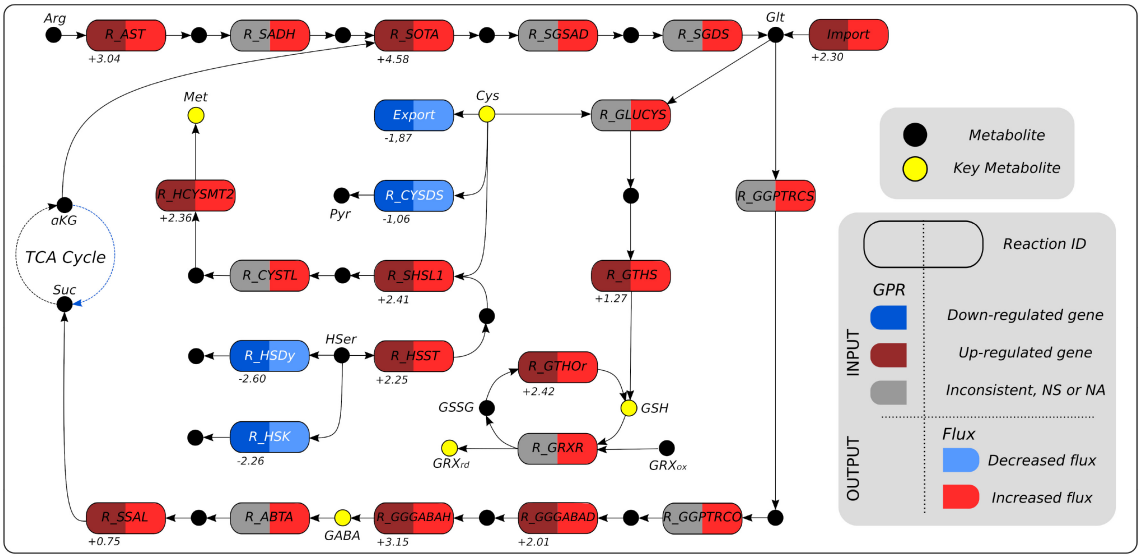

Fig. 3. Predicted flux changes for the production of key metabolites related to mercury stress response of *E.coli*. Key metabolites related to mercury stress response are highlighted in yellow. In the case of import of Glt, more than one reaction was considered, therefore, there are three different $\log_2$ fold changes reported. Grey left rectangles (INPUT) signify *a priori* grey reactions: DE results were either inconsistent, no associated gene exceeded threshold for DE (NS), or there were no known associated genes (NA). The metabolite abbreviations are defined in Supplementary Tables

the expected findings were inferred increases in reactions related to stress response and redox response (Fig. 3).

Reduced glutathione (GSH), a tripeptide composed of cysteine, glutamate and glycine, is considered to be one of the most important scavengers of reactive oxygen species (ROS), and is involved in the detoxification of certain xenobiotics and heavy metals. Its ratio with oxidized glutathione (GSSG) can be used as a marker of oxidative stress: in a resting cell, the molar GSH: GSSG ratio exceeds 100:1. However, it is altered under certain stress conditions when GSH is oxidized to GSSG (Zitka *et al.*, 2012). Following this, Lavoie and Summers expected to find an increase in cysteine and glutathione biosynthesis. However, most genes for biosynthesis of these biothiols appeared to be either down-regulated or showed no significant change.

`MOOMIN`, on the other hand, was able to detect these features, as shown in Figure 3. Even though there was no apparent increase in the production of cysteine (Cys), it is possible that more cysteine was available due to a decrease in its turnover to pyruvate (Pyr, R_CYSDS) and its export to the periplasm. We also predicted an increased production of glutamate (Glt) from arginine (Arg), which is essential for GSH biosynthesis.

Glutaredoxin (GRX) is also a key metabolite for redox homeostasis and stress response. Similarly to GSH, active GRX (GRXrd) is oxidized by substrates and it is only recycled to its active form with the concomitant oxidation of GSH (R_GRXR). This is why the increased flux towards the production of glutaredoxin (GRXrd) is closely linked to the recycling of GSSG into GSH (R_GTHor). MOOMIN also connected the down-regulation of genes responsible for the conversion of 2-oxoglutarate (αKG) to succinate (Suc) from the TCA cycle (blue arrow in Fig. 3) with the GABA shunt, something not described in LaVoie and Summers (2018) and also not apparent from the GO enrichment analysis. GABA (γ-aminobutyrate) is a metabolite that counteracts a wide range of stresses in several species, and seems to be essential for the acid tolerance in *E.coli* (Foster, 2004; Hersh *et al.*, 1996). We were also able to detect a possible increase in the fluxes from cysteine and homoserine (HSer) to another amino acid that has been linked to ROS scavenging in other organisms, namely methionine (Met) (Campbell *et al.*, 2016).

4.3 *Saccharomyces cerevisiae* ethanol stress

The final dataset we analyzed compared the response to ethanol stress under aerobic growth of two strains of *S.cerevisiae* with differing levels of ethanol tolerance: a tolerant strain Y7568 (PAP) and a susceptible strain NCYC3290 (WA). This data was originally published by Sardi *et al.* (2018) (GEO Accession ID: GSM3318043).

Although ethanol stress has been very well studied in the past decades in yeast, different levels of tolerance between strains are not yet completely understood. For instance, it has been reported that one of the main consequences of ethanol accumulation inside yeast cells results in ROS generation which, in turn, is responsible for inhibition of cell viability and cell death (Perez-Gallardo *et al.*, 2013).

Sardi *et al.* focussed on comparing the transcriptomic profiles of the strains stressed with ethanol by using the same strain under no stress as control. In the paper, the authors reported for all tested alcohols in the study (and in all strains) a general enrichment of genes involved in mitochondrial functions, siderophore transport, aryl-alcohol dehydrogenases and zinc binding.

Contrary to the analyses performed by Sardi *et al.*, we directly compared the stressed conditions: tolerant strain PAP after ethanol stress as the test condition against the susceptible strain WA after ethanol stress as the control condition. In this way, we did not aim at finding the reported general mechanisms of ethanol stress response, as most of them are shared between the strains, but rather factors that could influence the difference in tolerance. Similarly to what was reported, and in accordance with our GO enrichment analysis we were able to detect the decreased flux in fatty acid biosynthesis in the tolerant strain (28 out of 63 reactions in the pathway) and fatty acid degradation (29 out of 45 reactions). We also detected other interesting features that could be related to the difference in susceptibility. In addition to fatty acid metabolism, energy, alternate carbon, sterol and sphingolipid metabolisms, along with amino acid biosynthesis (with the exception of arginine), were predicted to have decreased fluxes in the tolerant strain, indicating that this strain might be able to shift its metabolism towards other pathways more important to alcohol defense. Some of these features were also detected by our GO enrichment analysis (Table 1). Although most of the genes involved in arginine biosynthesis were not detected as differentially expressed, MOOMIN successfully recovered a connected pathway with an increased flux towards arginine production (Fig. 4A). This is in agreement with previous studies that have shown the protective effects of arginine against ethanol stress in yeast (Cheng *et al.*, 2016), notably due to the direct protection against ROS generation. Indeed, this is an example of how the method is able to find new hypotheses not clear from only DE data and shows that incorporating the metabolic network structure into the analysis of transcriptomic data can help to connect isolated differentially expressed genes in a biologically meaningful way.

Another reason for the difference in susceptibility predicted by MOOMIN might be a higher availability of cytoplasmic trehalose. Several studies suggest that trehalose promotes the survival of yeast cells during ethanol stress, stabilizing or repairing proteins denatured by ethanol (Bandara *et al.*, 2009; Ding *et al.*, 2009). The transcriptomic data showed up-regulation of trehalose degradation in the cytoplasm (R_TREH) and down-regulation in vacuole (R_TREH_v). MOOMIN confirmed these changes, and connected them through the trehalose transport reaction (R_TREt2v) (Fig. 4B). Because the transport reaction is reversible, it was coloured yellow (see Section 3 and Fig. 1). Both interpretations—an increased transport from vacuole to the cytoplasm or a decreased transport in the opposite direction—are consistent with increased availability of trehalose in the cytoplasm. Down-regulation of vacuolar trehalase (related to the reaction R_TREHv) has already been shown to promote ethanol fermentation and tolerance (Jung and Park, 2005), whereas the cytosolic trehalase (related to the reaction R_TREH) was shown to be regulated by several types of stress (Thevelein, 1984).

In this way, MOOMIN was successful in predicting pathways that were previously implicated in ethanol tolerance even though some of these features were not significant and directly detectable from the DE data.

4.3.1 Comparison of the outputs from the full and the topological formulation

In order to better understand the differences between the outputs from the original problem [Equations (8)–(19)] and the topological approximation (see Supplementary Materials), we analyzed in detail both solutions for the *S.cerevisiae* resistance to ethanol dataset. Out of 1576 reactions analyzed in the network, 89% were in accordance between the topological results (consensus of 48 outputs) and the stoichiometric output (Supplementary Tables). Out of the 174 reactions with different outputs, 21 were exchange boundary reactions and 23 were extracellular transport, which means they resulted from different outcomes from within the network. From the remaining 130 intracellular reactions, 26 were considered inconsistent among the 48 topological solutions; this means that they were chosen at times as blue and red in at least two different outputs.

We further analyzed the 104 intracellular reactions: 57 were present only in the topological outputs and 47 were present only in the stoichiometric output. Strikingly, 40% of the reactions only present in the topological results were related to lipid metabolism (phospholipid, glycerolipid, fatty acid, sterol metabolisms). Indeed, there were two problems related to the stoichiometric balance of phospholipids in the nucleus that did not allow these 19 reactions to take place in the stoichiometric subnetwork: phosphatidyl-1D-myoinositol and ATP were detected as stoichiometric dead-end metabolites (DEM) in the nucleus. While four reactions consumed ATP in the nucleus, only one was able to recover it; however, the flux of recovery was not able to fully allow the flux in the other four reactions. This in turn blocked the following reactions related to these metabolites. Another 30% of the reactions included only in the topological results were related to nucleotide, cofactor and pentose phosphate pathways and resulted from the inclusion of ribose-1-phosphate, another stoichiometric DEM within the subnetwork. The remaining differences were mainly related to mitochondrial, nuclear and

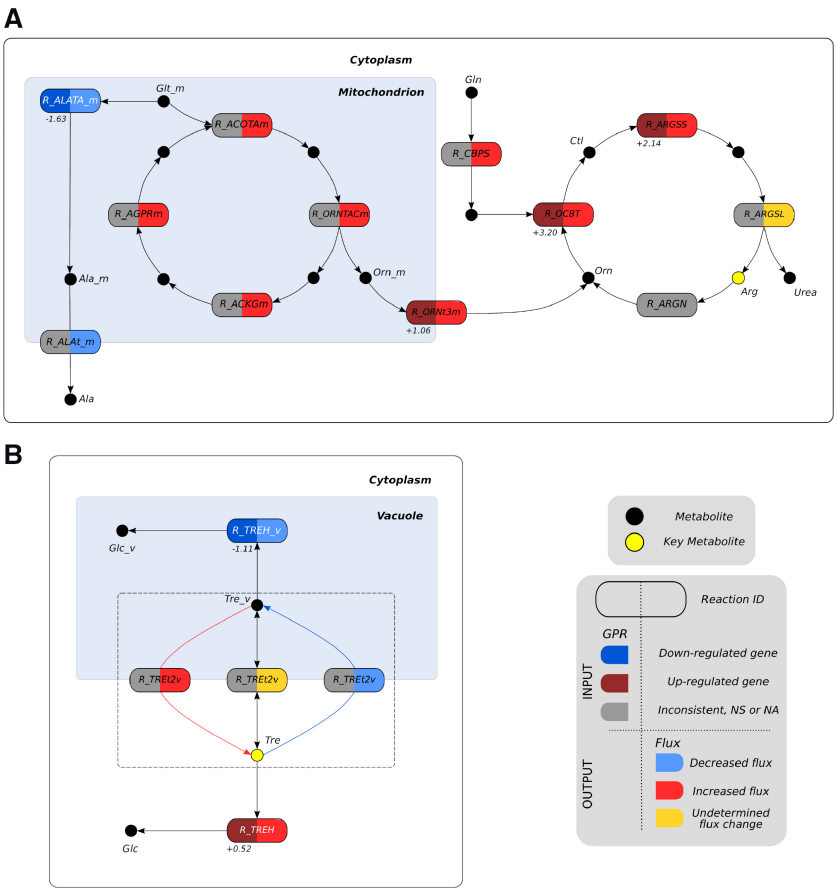

Fig. 4. Predicted flux changes for ethanol tolerance in *S.cerevisiae* in (**A**). arginine biosynthesis; (**B**) trehalose degradation pathway. The metabolite abbreviations are defined in Supplementary Tables

peroxisome reactions, due to the imbalance of several stoichiometric DEMs. Hence, the differences between solutions generally pointed to possible stoichiometric problems within the network.

5 Conclusion

In this paper, we have introduced a new computational method for the analysis of gene expression data that makes use of a genome-scale metabolic reconstruction of the organism under study: MOOMIN. The method takes as its input the results of a differential expression analysis comparing the gene expression between two different conditions and a metabolic network, and produces a hypothesis of a metabolic shift: a qualitative change in the activity of reactions indicating which parts of the metabolism increased, decreased or remained unchanged in flux between the two states. We have described the theoretical framework behind MOOMIN and its implementation in the language of Mixed-Integer Linear Programming.

We applied the MOOMIN method to three distinct datasets from both eukaryotic and prokaryotic organisms and were able to extract interesting information that were not possible to infer only by analyzing the DE data itself. In the latter two datasets, investigating *E.coli*'s response to mercury and *S.cerevisae*'s tolerance to ethanol, by analyzing at the same time the network information and the differentially expressed genes, MOOMIN inferred pathways that were already shown to be involved with the analyzed conditions but were not detected as differentially expressed by our GO enrichment analysis or by the authors from both publications, namely the GSH production in *E.coli* in response to mercury and the arginine production in strains of yeast resistant to ethanol.

The method was also able to infer pathways expected to be differentially expressed and that were also enriched in our GO analysis, but in a more connected way, focussing also on genes that possibly associate these seemingly unrelated functions and connecting up- and down-regulated genes. A good example of this feature is seen in Figure 3: MOOMIN recovered a connected subnetwork in which the TCA cycle and amino acid metabolism were directly linked to the production of GSH and GABA. Altogether, our analyses show that by coupling the network topology and DE data, MOOMIN is able to both further confirm results that can be obtained using traditional tools of DE analysis, as well as discover new information such as metabolically regulated reactions that cannot be straightforwardly inferred from gene expression data alone.

Acknowledgements

This work was performed using the computing facilities of the CC LBBE/PRABI.

Funding

This work was supported by the Horizon 2020 Program of the European Commission within the Marie Skłodowska-Curie Innovative Training Network MicroWine [grant number 643063]; the Netherlands Organisation for Scientific Research (NWO) through the Gravitation Programme Networks

[grant number 024.002.003] to L.S.; by the Agence Nationale de la Recherche (ANR) [grant number ANR-17-CE20-0031] to M.G.F.; and by the São Paulo Research Foundation (FAPESP) [grant number 2015/13430-9 and 2017/05986-2.] to R.A.

Conflict of Interest: none declared.

References

Agren,R. *et al*. (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol*., **8**, e1002518.

Akesson,M. *et al*. (2004) Integration of gene expression data into genome-scale metabolic models. *Metab. Eng*., **6**, 285–293.

Alcaraz,N. *et al*. (2011) Keypathwayminer: detecting case-specific biological pathways using expression data. *Internet Math*., **7**, 299–313.

Alcaraz,N. *et al*. (2012) Efficient key pathway mining: combining networks and OMICS data. *Integr. Biol*., **4**, 756–764.

Antonov,A.V. *et al*. (2008) KEGG spider: interpretation of genomics data in the context of the global gene metabolic network. *Genome Biol*., **9**, R179.

Bandara,A. *et al*. (2009) Trehalose promotes the survival of *Saccharomyces cerevisiae* during lethal ethanol stress, but does not influence growth under sublethal ethanol stress. *FEMS Yeast Res*., **9**, 1208–1216.

Baumbach,J. *et al*. (2012) Efficient algorithms for extracting biological key pathways with global constraints. In: *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, GECCO '12, pp. 169–176. ACM, New York, NY, USA.

Becker,S.A. and Palsson,B.O. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol*., **4**, e1000082.

Campbell,K. *et al*. (2016) Methionine metabolism alters oxidative stress resistance via the pentose phosphate pathway. *Antioxid. Redox Signal*., **24**, 543–547.

Chandrasekaran,S. and Price,N.D. (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA*, **107**, 17845–17850.

Cheng,Y. *et al*. (2016) Protective effects of arginine on *Saccharomyces cerevisiae* against ethanol stress. *Sci. Rep*., **6**, 31311.

Clarke,B.L. (1988) Stoichiometric network analysis. *Cell Biophys*., **12**, 237–253.

Colijn,C. *et al*. (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput. Biol*., **5**, e1000489.

Collins,S.B. *et al*. (2012) Temporal expression-based analysis of metabolism. *PLoS Comput. Biol*., **8**, e1002781.

Covert,M.W. and Palsson,B.O. (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol*., **221**, 309–325.

Ding,J. *et al*. (2009) Tolerance and stress response to ethanol in the yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol*., **85**, 253–263.

Fang,X. *et al*. (2012) Modeling phenotypic metabolic adaptations of *Mycobacterium tuberculosis* H37Rv under hypoxia. *PLoS Comput. Biol*., **8**, e1002688.

Foster,J.W. (2004) *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat. Rev. Microbiol*., **2**, 898–907.

Heirendt,L. *et al*. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc*., **14**, 639–702.

Hersh,B.M. *et al*. (1996) A glutamate-dependent acid resistance gene in *Escherichia coli*. *J. Bacteriol*., **178**, 3978–3981.

Ideker,T. *et al*. (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18** (**Suppl. 1**), S233–240.

Jensen,P.A. and Papin,J.A. (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, **27**, 541–547.

Jerby,L. *et al*. (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol*., **6**, 401.

Jung,Y.J. and Park,H.D. (2005) Antisense-mediated inhibition of acid trehalase (ATH1) gene expression promotes ethanol fermentation and tolerance in *Saccharomyces cerevisiae*. *Biotechnol. Lett*., **27**, 1855–1859.

Kim,H.U. *et al*. (2013) Flux-coupled genes and their use in metabolic flux analysis. *Biotechnol. J*., **8**, 1035–1042.

Kim,M.K. and Lun,D.S. (2014) Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput. Struct. Biotechnol. J*., **11**, 59–65.

Kim,M.K. *et al*. (2016) E-Flux2 and SPOT: validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. *PLoS One*, **11**, e0157101.

King,Z.A. *et al*. (2016) BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res*., **44**, D515–522.

Klamt,S. *et al*. (2009) Hypergraphs and cellular networks. *PLoS Comput. Biol*., **5**, e1000385.

Lacroix,V. *et al*. (2008) An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform*., **5**, 594–617.

LaVoie,S.P. and Summers,A.O. (2018) Correction to: transcriptional responses of *Escherichia coli* during recovery from inorganic or organic mercury exposure. *BMC Genomics*, **19**, 268.

Lee,D. *et al*. (2012) Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol*., **6**, 73.

Leng,N. *et al*. (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, **29**, 1035–1043.

Machado,D. and Herrgård,M. (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol*., **10**, e1003580.

Mo,M.L. *et al*. (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol*., **3**, 37.

Moxley,J.F. *et al*. (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc. Natl. Acad. Sci. USA*, **106**, 6477–6482.

Noirel,J. *et al*. (2008) Automated extraction of meaningful pathways from quantitative proteomics data. *Brief. Funct. Genomic Proteomic*, **7**, 136–146.

Nookaew,I. *et al*. (2012) A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res*., **40**, 10084–10097.

Opdam,S. *et al*. (2017) A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Syst*., **4**, 318–329.

Orth,J.D. *et al*. (2010) What is flux balance analysis? *Nat. Biotechnol*., **28**, 245–248.

Orth,J.D. *et al*. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Mol. Syst. Biol*., **7**, 535.

Perez-Gallardo,R.V. *et al*. (2013) Reactive oxygen species production induced by ethanol in *Saccharomyces cerevisiae* increases because of a dysfunctional mitochondrial iron-sulfur cluster assembly system. *FEMS Yeast Res*., **13**, 804–819.

Rezola,A. *et al*. (2015) Advances in network-based metabolic pathway analysis and gene expression data integration. *Brief. Bioinform*., **16**, 265–279.

Robaina Estevez,S. and Nikoloski,Z. (2015) Context-specific metabolic model extraction based on regularized least squares optimization. *PLoS One*, **10**, e0131875.

Rossell,S. *et al*. (2013) Inferring metabolic states in uncharacterized environments using gene-expression measurements. *PLoS Comput. Biol*., **9**, e1002988.

Saha,R. *et al*. (2014) Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol*., **29**, 39–45.

Samal,S.S. *et al*. (2017) Linking metabolic network features to phenotypes using sparse group lasso. *Bioinformatics*, **33**, 3445–3453.

Sardi,M. *et al*. (2018) Genotype-by-environment-by-environment interactions in the *Saccharomyces cerevisiae* transcriptomic response to alcohols and anaerobiosis. *G3 (Bethesda)*, **8**, 3881–3890.

Schultz,A. and Qutub,A.A. (2016) Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput. Biol*., **12**, e1004808.

Schwarz,R. *et al*. (2005) YANA – a software tool for analyzing flux modes, gene-expression and enzyme activities. *BMC Bioinformatics*, **6**, 135.

Segal,E. *et al*. (2003) Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics*, **19**, i264–271.

Shlomi,T. *et al*. (2008) Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol*., **26**, 1003–1010.

Thevelein,J.M. (1984) Regulation of trehalose mobilization in fungi. *Microbiol. Rev*., **48**, 42–59.

Tian,M. and Reed,J.L. (2018) Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics*, **34**, 3882–3888.

van Helden,J. *et al*. (2000) Representing and analysing molecular and cellular function using the computer. *Biol. Chem*., **381**, 921–935.

Vivek-Ananth,R.P. and Samal,A. (2016) Advances in the integration of transcriptional regulatory information into genome-scale metabolic models. *Biosystems*, **147**, 1–10.

Vlassis,N. *et al*. (2014) Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol*., **10**, e1003424.

Wang,Y. *et al*. (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol*., **6**, 153.

Yizhak,K. *et al*. (2014) Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *Elife*, **3**, e03641.

Zhang,S.W. *et al*. (2017) Prediction of metabolic fluxes from gene expression data with Huber penalty convex optimization function. *Mol. Biosyst*., **13**, 901–909.

Zhu,L. *et al*. (2017) A computational method using differential gene expression to predict altered metabolism of multicellular organisms. *Mol. Biosyst*., **13**, 2418–2427.

Zitka,O. *et al*. (2012) Redox status expressed as GSH: GSSG ratio as a marker for oxidative stress in paediatric tumour patients. *Oncol. Lett*., **4**, 1247–1253.

Zur,H. *et al*. (2010) iMAT: an integrative metabolic analysis tool. *Bioinformatics*, **26**, 3140–3142.